

文章编号: 1000-5013(2008)04-0518-04

# 利用 XML, Tamino 和 CORBA 的 软构件管理与检索技术

余金山

(华侨大学 信息科学与工程学院, 福建 泉州 362021)

**摘要:** 给出一种基于 XML, Tamino 和 CORBA 的软构件库管理与检索的方案和架构体系。它是一个松耦合的 3 层结构, 数据库服务器负责管理和存储所有 XML 文档和 Schema 文件, 是由 Tamino 提供支持的, 主要包括 X-Machine、数据映射、XML 解析器、对象处理器、查询解释器和对象合成器等 6 个功能模块。已实现的实验性系统具有数据加载、构件管理、用户管理、构件检索等主要功能。

**关键词:** 软构件; 构件库管理系统; XML; Tamino

**中图分类号:** TP 311.5

**文献标识码:** A

基于构件的软件开发(CBSD)或称为基于构件的软件工程(CBSE), 被认为是一种革命性的软件开发新模式<sup>[1]</sup>。现有的研究不能有效、充分地发现和利用构件源。完整的构件库是实现 CBSD 的关键, 而各个开发者或开发机构要创建各自的、绝对完整的构件库是不可能的。因此, 为各种异构的、分布的、多样化的和各种形式的构件库提供一个较好的联合、协作和共享机制, 并为用户提供一种统一而且简单有效的手段, 以便从最丰富的构件源中检索/获取最能满足自己需求的构件, 是目前一个研究的重点<sup>[2-11]</sup>。本文给出一个基于 XML(Extensible Markup Language), Tamino 和公共对象请求代理体系结构(Common Object Request Broker Architecture, CORBA)的软构件管理技术, 以期能较好地解决这一问题。

## 1 解决方案

### 1.1 涉及的子问题

首先, 应该考虑如下 5 个方面的内容。

(1) 构件库之间的协作共享必须有一个公共的管理平台, 以解决分布、异构、信任、知识产权、安全性、商业和市场因素等问题。此外, 还要提供一种构件库之间可用最简单、最省力的方式进行灵活的、自由的协作, 又能充分共享的技术和方法。

(2) 访问构件库的用户可能具有不同的层次和不同的教育文化背景。应提供支持构件使用者用自己最熟悉的或最喜欢的表达方式进行多库协同检索的描述手段和工具, 从而使用户能用最简单的方式, 从最丰富的构件源中获取满足自己需求的构件。

(3) 不影响或尽量少影响不同的机构或个人建立自成体系的、有自己特色的、能提供多样化分类和检索模式的构件系统。构件的多分类多检索模式是几乎所有研究公认应该支持的。

(4) 能对剖面、属性-值、关键词、枚举、正文、形式化等多种分类和检索方法提供统一的描述。同时, 有利于与人工智能技术(AI)结合以处理自然语言查询, 为用户提供更有效、更全面的检索机制。

(5) 提供对代码级、设计级、分析级、测试级等各抽象层次的复用的支持, 包括对需求规约、设计、模式、测试计划等文档知识的构件复用。随着万维网和 Internet/ Intranet 的迅速发展, 已经出现了许多新

收稿日期: 2008-07-20

作者简介: 余金山(1952-), 男, 教授, 主要从事为软件工程和人工智能应用的研究。E-mail: yjs@hqu.edu.cn.

基金项目: 福建省自然科学基金资助项目(A0810013)

形式的构件, 如电子邮件、HTML 文档、音频、图像、动画、甚至视频等。异构构件源的集成与共享, 除了要集成传统的构件库外, 还要能管理和集成这些新的构件源。

1.2 相关技术的特点

XML 本身所具有的特点及 Web 应用需求的推动, 使得 XML 迅速地超越了其用于 Web 上的文档传输的设计初衷, 而成为 Web 上跨平台的信息表示、信息交换和互操作的标准, 并逐渐成为建立新的存储技术的基础<sup>[12-14]</sup>。XML 可以为构件的使用者在需用访问不同的构件源时提供标准的统一接口, 使得构件的生产者可以建立自成体系的、有自己特色的、能提供多样化分类和检索模式的构件系统, 同时也使得第三方的服务商不用被局限于特定的工具和环境。

XML 文档的形式存在, 采用基于 XML 数据库管理系统来完成构件库创建。目前基于 XML 的数据库主要分为两类: XED(XML-Enabled Database 或 XML-Enabled DBMS) 和 NXD(Native XML Database 或 Native XML DBMS)。XED 和 NXD 各有其优缺点<sup>[15-16]</sup>。XML 文档格式复杂, 处理的又是多种数据格式和大文档数据, 数据本身有层次性关系, 而且要求数据描述能力高, 可方便在网络上进行交换处理等; 而 NXD 能提供更好的性能和更完备的支持。

针对 NXD 的这些优点, 选用 Tamino 来完成构件库的创建和构件描述文档的控制。Tamino 本质上是一个使用纯粹且标准的 XML 格式, 可进行存储和获取各种类型数据的信息服务器。它提供了直接保存 XML 数据的方案, 采用层次数据存储模型, 保持 XML 文档的树形结构。软构件描述本身就有层次结构, 例如构件分类、剖面描述等。Tamino 实现了完整的 XML 数据库系统, 提供了完善的 XML 文档数据的管理与检索功能和各种标准数据访问接口。但是, 为实现面向网络的跨平台、跨语言、跨地域、跨部门和跨多个不同分类模式的构件库系统之间的检索机制, 为用户有效屏蔽软、硬件环境和不同构件分类方案间的差异。目前最好的方法是采用中间件技术。CORBA 技术作为一种成熟的中间件技术, 具有模型完整、先进性, 以及独立于系统平台和开发语言, 可集成已有系统并屏蔽通信和实现细节等优点。

2 体系结构

基于 XML, Tamino 和 CORBA 技术的软构件系统的体系结构, 如图 1 所示。体系是一个松耦合的

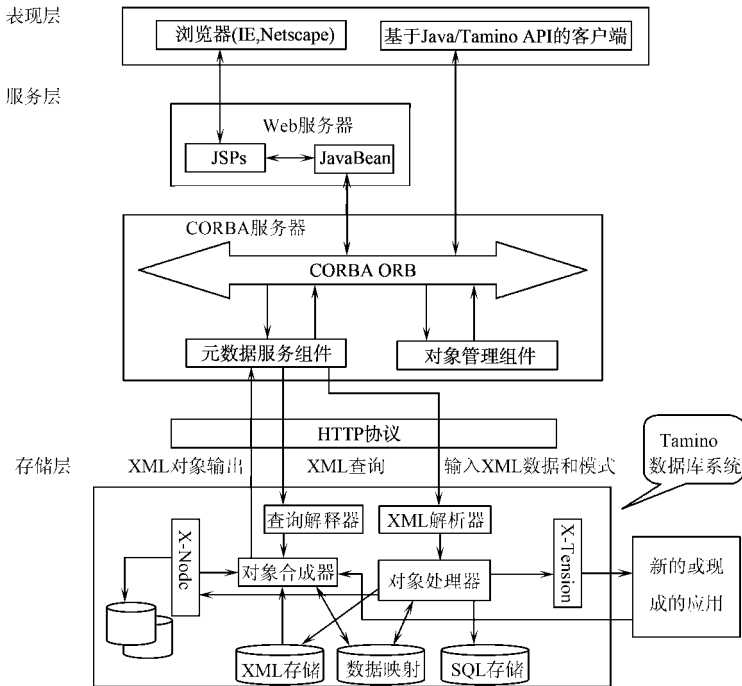


图 1 软构件系统的体系结构

Fig. 1 Architecture of software components system

3 层结构, 可分为表现层、服务层和存储层。表现层是基于浏览器或者 Java 客户应用程序。它通过 XML 语言为不同的客户提供统一的获取分布式异构环境下的软构件信息和服务, 并为构件的管理者提供基

于 XML 的构件描述和管理工具. 服务层是通过在 CORBA 服务器和 Web 服务器中部署主要的管理和 服务组件, 包括检索管理、元数据服务、对象管理服务和安全管理组件, 提供相应的认证、调度、发布及检索功能.

数据库服务器负责管理和存储所有 XML 文档和 Schema 文件, 是由 Tamino 提供支持的, 主要包括如下 6 个功能模块.

(1) X-Machine 驱动 Tamino 的引擎, 基本功能是存储 XML 对象并从不同的数据源中进行查询. 它是基于管理员定义在数据映射中的模式来处理存储和查询的. XML 对象存储在 Tamino 中, 但是 Tamino 也提供一个内部的 SQL 存储, 并支持 Tamino 以外的其他数据库类型.

(2) 数据映射. Tamino 的知识库, 主要由管理员制定, 包含存储和查询及组织生成 XML 对象的规划. 这些模式的制定可以由图形工具来支持并保持其模式树的正确性.

(3) XML 解析器. 存储在 X-Machine 的 XML 对象, 由存储在 Tamino 数据映射中的模式进行描述. X-Machine 的 XML 解析器会检查模式的语法正确性, 以确保加载的 XML 对象是“格式良好”的.

(4) 对象处理器. Tamino 存储数据时, 对象处理器使用数据映射中提供的相关模式信息把 XML 数据存储在 XML 存储中.

(5) 查询解释器. Tamino 的查询语言是 X-Query. 查询解释器分解查询请求并与对象合成器交互, 按照存储在数据映射中的模式检索对象.

(6) 对象合成器. 使用定义在数据映射中的存储和查询规则, 对象合成器能够构造信息对象并把它们作为 XML 文档返回.

### 3 应用系统的主要功能

根据以上所述, 已实现的实验性系统包括如下 4 个功能.

(1) 数据加载. 提供基于 XML 的构件描述和管理工具. 根据基于 XML 的构件描述 Schema, 将构件信息和术语信息转换成 XML 格式文档(由服务层元数据服务组件实现), 并通过 Tamino 解析器来解析文档模式的语法正确性. 然后, 与数据映射中提供的查询和存储的相应规则一起提交给对象处理器, 将转换后的 XML 数据存储到 Tamino 数据库中. 同时, 利用 Tamino 提供的索引引擎, 为构件描述文档建立检索索引库.

(2) 构件管理. 提供一个包括构件的查询、修改、删除、备份、恢复、访问权限控制和日志管理在内的管理工具. 其中, 数据查询提供标准的 XML 查询语言 X-Query; 管理人员能对构件库中以树型结构保存的数据进行灵活的、深入到不同节点执行查询、修改、删除、备份、恢复、访问权限控制等管理.

(3) 用户管理. 为系统管理员提供和多级用户访问权限控制的管理机制.

(4) 构件检索. 提供各种检索机制, 如全文检索、关键词检索、复杂单元检索, 并将全文检索转换成基于剖面树模式的剖面检索.

用户在客户端输入查询条件后, 由服务层中的元数据服务组件负责根据用户查询条件生成 X-Query 查询语句. 根据查询条件的剖面术语检索 Tamino 数据库中的词典文档获取各术语相应的同义词, 并将术语与同义词作为 X-Query 语句中的查询条件. 将条件中的剖面术语与检索出的近义词作为构件剖面查询条件值序列组成 X-Query 语句, 然后将生成的 X-Query 查询语句通过 Tamino 数据访问接口发送到 Tamino 数据库管理系统, 查询解释器分解查询请求. 按照数据映射提供的存储规则与对象合成器交互, 使用 XML 数据的无序树匹配技术, 检索 XML 数据库并返回查询结果.

### 4 结束语

给出的架构和技术不但可用于多库的协作和共享, 也适用于内部的或单一库的管理与检索. 此外, 它还将有利于把语义 Web, Internet 语义信息处理技术、网络计算平台和架构技术、知识管理技术等融入到构件技术中, 从而更好地解决软构件的分类、描述和检索问题. 基于 XML, NXD 和中间件的软构件库架构将有较好的应用前景.

## 参考文献:

- [1] 余金山, 马启波. 一种基于 Web 的分布式构件库组织框架[J]. 计算机科学, 2002, 29(5): 152-154.
- [2] 潘 颖, 赵俊峰, 谢 冰, 等. 构件库技术的研究与发展[J]. 计算机科学, 2003, 30(5): 90-93.
- [3] DEAN J, OBENDORF P. COTS workshop: Continuing collaboration for successful COTS development[J]. Software Engineering Notes, 2001, 26(1): 61-73.
- [4] BACHMAN F, BASS L, BUHMAN C. Technical concepts of component-based software engineering[R]. USA: CMU/SEI-2000-TR-008, 2000.
- [5] SEACPRD R C, HISSAM S A, WALLNAU K C. Agoro—a search engine for component[J]. IEEE Internet Computing, 1998(11): 62-70.
- [6] GUO J, LU Q. A survey of software reuse repositories[C] // The IEEE Intl Conf and Workshop on the Engineering of Computer Based Systems Edinburgh: [s. n.], 2000.
- [7] 徐如志, 钱乐秋, 程建平, 等. 基于 XML 的软构件查询匹配算法研究[J]. 软件学报, 2003, 14(7): 1195-1202.
- [8] WING J M. Specification matching of software components[J]. ACM Transaction on Software Engineering and Methodology, 1997, 16(4): 333-369.
- [9] GIBB F, MCCARTAN C, O'DONNELL R, et al. The integration of information retrieval techniques within a software reuse environment[J]. Journal of Information Science, 2000, 26(4): 520-539.
- [10] TORSHEEN S. Design and implementation of an approximate pattern matching language for XML[R]. Berlin: Berlin Freie University, 2001.
- [11] 高 明, 陈 昕, 李 炜, 等. 基于 XML 实现异构数据源的联合使用[J]. 计算机科学, 2002, 29(3): 83-84.
- [12] 李立宇, 谭少华, 唐世渭, 等. 语义 Web 及其描述语言[J]. 计算机科学, 2003, 30(12): 1-5.
- [13] 司功闪, 王鸿谷, 徐 捷. 以 XML 为核心的 Tamino 数据库的研究与分析[J]. 计算机工程, 2004, 30(16): 78-79.
- [14] 秦 杰, 杨树强, 冀文华. XML 数据库技术研究[J]. 计算机科学, 2003, 30(9): 6-9.
- [15] 王照岳, 孙建伶, 董金祥. XML 数据库管理系统研究[J]. 计算机科学, 2002, 29(1): 115-117.
- [16] 王黎敏, 陈 奇, 俞瑞钊. XML 与 CORBA 的结合: 提高分布式系统的可扩展性[J]. 计算机科学, 2002, 29(5): 4-8.

## Based on XML-Tamino-CORBA Technology for Software Components Management and Retrieval

YU Jin-shan

(College of Information Science and Engineering, Huaqiao University, Quanzhou 362021, China)

**Abstract:** A XML-Tamino-CORBA based scheme and architecture is presented for software components management and retrieval, which is a loosely coupled three-tier architecture. Database server takes charge of management and storage of all XML documents and Schema files, which is supported by Tamino and comprized X-Machine, data map, XML parse, object processor, query interpreter and object combiner. Realized experimental system has data-loading, component management, user management, component retrieval function and so on.

**Keywords:** software components; components management system; XML; tamino

(责任编辑: 鲁 斌 英文审校: 吴逢铁)